NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2000; **00**:1–6 *Prepared using* nlaauth.cls *[Version: 2002/09/18 v1.02]*

On the two-grid convergence estimates

Robert D. Falgout[1], Panayot S. Vassilevski[1,*] and Ludmil T. Zikatanov[2]

[1] *Center for Applied Scientific Computing, UC Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551, U.S.A., Email: rfalgout@llnl.gov, panayot@llnl.gov*
[2] *Center for Computational Mathematics and Applications, Department of Mathematics, The Pennsylvania State University, University Park, PA 16802 Email: ludmil@psu.edu*

SUMMARY

We derive a new representation for the exact convergence factor of the classical two-level and two-grid preconditioners. Based on this result, we establish necessary and sufficient conditions for constructing the components of efficient algebraic multigrid (AMG) methods. The relation of the sharp estimate to the classical two-level hierarchical basis methods is discussed as well. Lastly, as an application, we give an optimal two-grid convergence proof of a purely algebraic "window"-AMG method. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: two-grid, two–level methods, convergence, sharp estimates, algebraic multigrid

1. Introduction

In this paper we are concerned with two-level hierarchical basis (TL) or more-generally with two-grid (TG) methods. We consider a vector space V, isomorphic to a $\mathbb{R}^n$ for some n. The space V is equipped with the usual Euclidean vector inner product $\langle \mathbf{v},\ \mathbf{w}\rangle = \mathbf{w}^T\mathbf{v}$. Our focus will be on two-level or two-grid iterative methods, for the solution of

$$A\mathbf{u} = \mathbf{f}, \tag{1.1}$$

where $A : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a *symmetric* and *positive definite* (s.p.d.) matrix. A linear iterative method takes the form: Given an initial guess $\mathbf{u}^{(0)}$, we obtain each successive iterate as

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + B^{-1}(\mathbf{f} - A\mathbf{u}^{(k)}). \tag{1.2}$$

---

*Correspondence to: Panayot S. Vassilevski, Center for Applied Scientific Computing, UC Lawrence Livermore National Laboratory, P.O. Box 808, L-560, Livermore, CA 94551, U.S.A.

Contract/grant sponsor: The work of the third author was supported in part by the National Science Foundation; contract/grant number: DMS-0209497

Contract/grant sponsor: This work was performed under the auspices of the U. S. Department of Energy by University of California Lawrence Livermore National Laboratory; contract/grant number: W-7405-Eng-48

Here we consider the case when the *iterator* or *preconditioner* B is defined via two-level or two-grid algorithms. Of main interest is the convergence of the approximate solutions $\mathbf{u}^{(k)}$ defined via (1.2) to the solution $\mathbf{u}$ of (1.1). This convergence rate is determined by estimating the norm defined by the A-inner product $(.)^T A(.)$ (or the A-norm $\|\cdot\|_A$) of the error transfer operator $E = I - B^{-1}A$. Whenever needed, we will distinguish between two-level hierarchical basis (HB) methods and two-grid methods by denoting the corresponding preconditioner B and the error transfer operator E with B_{TL}, E_{TL} and B_{TG}, E_{TG} respectively.

For two-level hierarchical basis methods, we assume that V is decomposed as a *direct sum*

$$V = SV_s + R^T V_c, \tag{1.3}$$

for some components V_s and V_c isomorphic to $\mathbb{R}^{n_s}$ and $\mathbb{R}^{n_c}$ respectively, with $n = n_s + n_c$. A typical and simple example to keep in mind is $S = [\ I \ \ 0\ ]^T$ and $R^T = [\ 0 \ \ I\ ]^T$. Here, the decomposition is actually orthogonal, i.e., $RS = 0$. In general, for a direct decomposition, we have that the square matrix $[S,\ R^T]$ is invertible. For the more general case of "two-grid" methods, one also assumes a decomposition $V = SV_s + R^T V_c$ as in (1.3), but it does not have to be direct. That is, one may have $n \leq n_s + n_c$. Yet another difference is that for two-grid methods only the *coarse* space $R^T V_c$ is explicitly available (i.e., a basis is given), whereas the first component (which is not unique) need not actually be specified. Usually, a one to one mapping P defines the coarse space. Such a mapping is called *interpolation* or *prolongation*, which in practice is sought to be in a sense better than R^T. One has $PV_c \subset V$, and the coarse space is Range(P). In the two-level hierarchical basis case, we assume that the decomposition $V = SV_s + PV_c$ is also a direct sum (this decomposition is commonly referred to as the two-level hierarchical basis decomposition). For example, in the simple case when $S = [\ I \ \ 0\ ]^T$ and $R^T = [\ 0 \ \ I\ ]^T$, the decomposition is direct if $RP = I$, i.e., if $P = \left[\ W^T \ \ I\ \right]^T$ for some $W \neq 0$. Note that the property, $RP = I$, often holds in the case of the more standard two-grid methods as well.

1.1. Smoother, coarse grid matrix and projections

A two-level or two-grid method can be defined if two ingredients are in place. One of them is the decomposition and the other is the *smoother*. A smoother here will be denoted with M, where the smoother iteration is as in (1.2) with M replacing B. The following result is well known (and easily seen):

$$M^T + M - A \text{ is s.p.d.} \quad \Longleftrightarrow \quad \|I - M^{-1}A\|_A < 1. \tag{1.4}$$

Hence, throughout the remainder of the paper we will assume that $M^T + M - A$ is s.p.d., or equivalently, that the smoother iteration is a contraction in A-norm.

Various restrictions of M and A to the subspaces mentioned before will be needed. We first define the exact coarse grid matrix A_c and its hierarchical complement A_s as follows

$$A_c = P^T AP, \qquad A_s = S^T AS.$$

Later we will see, in the case of a two level hierarchical preconditioner, one needs M to be well-defined only on the first (hierarchical) component SV_s. In that case, we refer to M as M_s. However, we can think of M_s as being derived from a global (not necessarily symmetric) smoother M, i.e., that $M_s = S^T MS$ where $M^T + M - A$ is positive semi-definite. As an example, consider again the simple case where $S = [\ I \ \ 0\ ]^T$ and $R^T = [\ 0 \ \ I\ ]^T$. Then, for a given M_s such that $M_s^T + M_s - A_s$ is positive definite, of interest is the block-factored smoother

$$M = \begin{bmatrix} M_s & 0 \\ RAS & \tau\, I \end{bmatrix} \begin{bmatrix} I & M_s^{-1}\, S^T A R^T \\ 0 & I \end{bmatrix},$$

where $\tau > 0$ is a sufficiently large constant. Note that (since $A^T = A$)

$$M^T + M - A = \begin{bmatrix} M_s^T + M_s - A_s & S^T A R^T \\ RAS & 2\tau I - RAR^T + (RAS)\left(M_s^{-1} + M_s^{-T}\right)(S^T A R^T) \end{bmatrix},$$

which can be made positive definite if τ is sufficiently large.

The following two operators are related to the smoother M, and will be frequently used in the definitions and analysis later on:

$$\widetilde{M} = M^T (M^T + M - A)^{-1} M, \qquad \overline{M} = M (M^T + M - A)^{-1} M^T. \tag{1.5}$$

Likewise, the operators $\widetilde{M}_s$ and $\overline{M}_s$ are defined by replacing M and A in (1.5) by M_s and A_s. Note that $(I - \widetilde{M}^{-1} A) = (I - M^{-1} A)(I - M^{-T} A)$, hence $\widetilde{M}$ is just a symmetrized version of the smoother M (and similarly for $\overline{M}$). Also note that $\widetilde{M} = \overline{M}$ when M is symmetric, but in general both operators are needed: $\widetilde{M}$ is needed for the error analysis and $\overline{M}$ is needed in the definition of the preconditioners. Finally, we remark that if $M^T + M - A$ is positive definite, then $\widetilde{M} - A$ and $\overline{M} - A$ are positive semidefinite. This is easily seen from the simple relation,

$$\widetilde{M} - A = (X - M) X^{-1} (X - M^T), \text{ with } X = M^T + M - A, \tag{1.6}$$

which, with obvious change, holds for $\overline{M}$ as well.

In what follows we will need two projection operators related to the coarse space Range(P). We define

$$\pi_A = P A_c^{-1} P^T A, \qquad \overline{\pi}_A = A^{\frac{1}{2}} P A_c^{-1} P^T A^{\frac{1}{2}}, \tag{1.7}$$

and observe that π_A is an A-orthogonal projection on Range(P) and $\overline{\pi}_A$ is a $\langle \cdot, \cdot \rangle$-orthogonal projection on Range($A^{\frac{1}{2}} P$).

1.2. The strengthened Cauchy-Schwarz inequality and the Schur complement

In the analysis of the two-level hierarchical preconditioner, we will need the strengthened Cauchy-Schwarz inequality (sometimes called the Cauchy-Bunyakowski-Schwarz, or C.B.S. inequality), which provides a bound on the cosine of the abstract angle between two subspaces. Assume that $V = SV_s + PV_c$ is a direct decomposition and let $\gamma^2 \in [0, 1)$ be the smallest constant in the following inequality

$$(\mathbf{w}^T S^T A P \mathbf{x})^2 \leq \gamma^2\; \mathbf{w}^T S^T A S \mathbf{w}\; \mathbf{x}^T P^T A P \mathbf{x}. \tag{1.8}$$

An equivalent form of this C.B.S. inequality reads

$$\mathbf{w}^T A_s \mathbf{w} \leq \frac{1}{1 - \gamma^2} \inf_{\mathbf{x}} \left(S\mathbf{w} + P\mathbf{x}\right)^T A \left(S\mathbf{w} + P\mathbf{x}\right), \quad \forall \mathbf{w} \in V_s, \quad \forall \mathbf{x} \in V_c.$$

The latter minimum is attained at $P\mathbf{x} = -\pi_A S \mathbf{w}$. Therefore,

$$\begin{aligned} \mathbf{w}^T A_s \mathbf{w} &\leq \tfrac{1}{1-\gamma^2}\, \mathbf{w}^T \left(S^T (I - \pi_A)^T A (I - \pi_A) S\right) \mathbf{w} \\ &= \tfrac{1}{1-\gamma^2}\, \mathbf{w}^T \left(S^T A (I - \pi_A) S\right) \mathbf{w}. \end{aligned} \tag{1.9}$$

As it is well known, the constant in the strengthened C.B.S. inequality is related to the spectral equivalence between the Schur complement $\mathcal{S}_A$ of A and $A_c = P^T AP$. The Schur complement $\mathcal{S}_A$ is defined here as

$$\mathbf{x}^T \mathcal{S}_A \mathbf{x} = \inf_{\mathbf{w}} \, (S\mathbf{w} + P\mathbf{x})^T A(S\mathbf{w} + P\mathbf{x}). \tag{1.10}$$

An equivalent statement of the strengthened Cauchy-Schwarz inequality (1.8) then is as follows

$$(1-\gamma^2)\, \mathbf{x}^T A_c \mathbf{x} \le \mathbf{x}^T \mathcal{S}_A \mathbf{x} \le \mathbf{x}^T A_c \mathbf{x}. \tag{1.11}$$

1.3. Two-level and two-grid preconditioners

Having all components in place, we are now in a position to define the classical two-level HB method (cf. Bank and Dupont [3], Braess [6], and Axelsson and Gustafsson [2], and [15], [4]). Most of these methods are summarized in Bank [5], (see also [10] or [13]). The two-level method in question, with parameters M_s and P, exploits a direct (hierarchical) decomposition $V = SV_s + PV_c$. Namely, we decompose $\mathbf{u} \in V$ uniquely as $\mathbf{u} = S\mathbf{u}_s + P\mathbf{u}_c$. The problem (1.1) is then transformed to the equivalent one, with the hierarchical basis matrix $\widehat{A} \equiv [S,\ P]^T A[S,\ P]$:

$$[S,\ P]^T A[S,\ P] \begin{bmatrix} \mathbf{u}_s \\ \mathbf{u}_c \end{bmatrix} = [S,\ P]^T \mathbf{f}.$$

This transformed problem is then used to define the preconditioner B_{TL} in terms of its hierarchical counterpart $\widehat{B}_{TL}$.

Definition 1.1 (Two-level hierarchical basis preconditioner, B_{TL}) *Let*

$$\widehat{B}_{TL} = \begin{bmatrix} I & 0 \\ P^T ASM_s^{-1} & I \end{bmatrix} \begin{bmatrix} \overline{M}_s & 0 \\ 0 & A_c \end{bmatrix} \begin{bmatrix} I & M_s^{-T} S^T AP \\ 0 & I \end{bmatrix}.$$

Then, the two-level hierarchical basis preconditioner B_{TL} is defined by

$$B_{TL}^{-1} = [S,\ P]\widehat{B}_{TL}^{-1}[S,\ P]^T.$$

It is clear, that to implement the actions of B_{TL}^{-1}, one needs the actions of the smoothers M_s^{-1}, M_s^{-T}, a coarse-grid solver to evaluate A_c^{-1}, the actions of the interpolation (P) and restriction (P^T) mappings, as well as the ability to extract the hierarchical component $S\mathbf{v}_s$ of any vector $\mathbf{v} = S\mathbf{v}_s + R^T\mathbf{v}_c$, which defines S and similarly for S^T. We comment on the fact that the above definition takes the point of view that B_{TL} is being obtained via a *product iteration scheme*, in the sense that the error transfer operator E_{TL} has the following form

$$E_{TL} \equiv (I - SM_s^{-T} S^T A)(I - PA_c^{-1} P^T A)(I - SM_s^{-1} S^T A).$$

The latter represents a composite subspace iteration process; the first step being smoothing based on M_s in the first coordinate space SV_s, followed by an (exact) coarse-grid correction in the subspace PV_c and finally followed by a post-smoothing based on M_s^T in the first coordinate space SV_s. From the preconditioning point of view though one can simplify the definition somewhat and thus end up with the following "approximate block-factorization" preconditioner in the case of s.p.d. M_s,

$$\widehat{B}_{TL}^{a} = \begin{bmatrix} M_s & 0 \\ P^T AS & I \end{bmatrix} \begin{bmatrix} M_s^{-1} & 0 \\ 0 & B_c \end{bmatrix} \begin{bmatrix} M_s & S^T AP \\ 0 & I \end{bmatrix}. \tag{1.12}$$

Then, $(B_{TL}^a)^{-1} = [S,\ P](\widehat{B}_{TL}^a)^{-1}[S,\ P]^T$ is the inverse "approximate block-factorization" preconditioner to A. This was the preconditioner originally studied in Axelsson and Gustafsson [2] with B_c being an s.p.d. approximation to A_c.

An important observation is that in order to define $B_{TL}^{-1} = [S,\ P]\widehat{B}_{TL}^{-1}[S,\ P]^T$, one does not have to assume that $[S,\ P]$ is invertible, or even square. Thus, one can formally let $S = I$, and hence $M_s = M$, and the resulting method gives the classical two-grid preconditioner B_{TG} with the corresponding error transfer operator defined by,

$$E_{TG} \equiv (I - M^{-T}A)(I - PA_c^{-1}P^TA)(I - M^{-1}A).$$

A more precise definition is as follows.

Definition 1.2 (Two-grid preconditioner, B_{TG}) *Let*

$$\widehat{B}_{TG} = \begin{bmatrix} I & 0 \\ P^TAM^{-1} & I \end{bmatrix} \begin{bmatrix} \overline{M} & 0 \\ 0 & A_c \end{bmatrix} \begin{bmatrix} I & M^{-T}AP \\ 0 & I \end{bmatrix}. \tag{1.13}$$

Then, the two-grid preconditioner B_{TG} is defined by

$$B_{TG}^{-1} = [I,\ P]\,\widehat{B}_{TG}^{-1}\,[I,\ P]^T. \tag{1.14}$$

Then it is easy to verify that

$$I - B_{TG}^{-1}A = E_{TG}.$$

Note that $\widehat{B}_{TG} :\ \mathbb{R}^{n+n_c} \mapsto \mathbb{R}^{n+n_c}$ has a "bigger" size than B_{TG} and A; namely, it defines an operator acting on the product space $V \times \text{Range}(P)$.

¿From (1.13), a straightforward calculation of the inverse B_{TG}^{-1}, gives

$$\widehat{B}_{TG}^{-1} = \begin{bmatrix} I & -M^{-T}AP \\ 0 & I \end{bmatrix} \begin{bmatrix} \overline{M}^{-1} & 0 \\ 0 & A_c^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -P^TAM^{-1} & I \end{bmatrix}. \tag{1.15}$$

Then, forming the right-hand side of (1.14) leads to

$$\begin{aligned} B_{TG}^{-1} &= \left[I,\ \left(I - M^{-T}A\right)P\right] \begin{bmatrix} \overline{M}^{-1} & 0 \\ 0 & A_c^{-1} \end{bmatrix} \begin{bmatrix} I \\ P^T\left(I - AM^{-1}\right) \end{bmatrix} \\ &= \overline{M}^{-1} + \left(I - M^{-T}A\right)PA_c^{-1}P^T\left(I - AM^{-1}\right). \end{aligned}$$

1.4. Main goals and structure of the paper

The last two definitions and the identities that follow are the main ingredients for comparing some classical convergence results with the new ones (derived in the present paper) as well as finding necessary and sufficient conditions for P and M (or M_s) to give optimal two-level or two-grid convergence in the related two-level or two-grid methods. This is the main topic of the present paper.

The remainder of the paper is structured as follows. In Section 2, we prove an important auxiliary result, the so-called "saddle-point" lemma. In Section 3, we use this lemma to prove the main (sharp) two-grid convergence result: a simple identity for the exact two-grid convergence factor. The result can also be derived from the identity of Xu and Zikatanov [14] (valid for abstract iteration methods), but we provide a more direct proof here. In Section 4, we relate the sharp two-grid convergence result with the existing tools commonly utilized for proving two-grid convergence in algebraic multigrid (AMG) theory. Many facts in Section 4 are simply reformulations of results already established in a previous paper [11], but here we are able to establish some additional necessary conditions. Finally, in Section 5 we show how the presented sharp convergence result can be used to derive upper bounds for the two-grid convergence rate of a so-called "window"-based spectral AMG method (a variant of the method proposed in Chartier et al. [9]).

2. A saddle-point lemma

A crucial identity that will be used to derive the spectral equivalence results is the following lemma, henceforth referred to as the "saddle-point lemma". We state the lemma in a somewhat abstract form for two vector spaces V_1 and V_2. We will use it with $V_1 = V_s$ for the two-level hierarchical basis preconditioner, and with $V_1 = V$ for the two grid preconditioner. The second space V_2 will be taken to be the range of a projection on a subspace of V.

Lemma 2.1. *Given two mappings $T\ : V_1 \mapsto V_1$ and $N :\ V_1 \mapsto V_2$ such that $T + N^T N$ is invertible, with T symmetric positive semi-definite and N onto (i.e., for any vector $\mathbf{v} \in V_2$ the equation $N\mathbf{w} = \mathbf{v}$ has at least one solution $\mathbf{w} \in V_1$). Consider the mapping $Z = N\left(T + N^T N\right)^{-1} N^T$. We have that Z is s.p.d., and the following identity holds:*

$$\frac{\mathbf{v}^T Z^{-1} \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = 1 + \inf_{\mathbf{w}:\ N\mathbf{w}=\mathbf{v}} \frac{\mathbf{w}^T T \mathbf{w}}{\mathbf{v}^T \mathbf{v}}. \tag{2.1}$$

Proof. We first remark that N being onto is equivalent to N^T having full column rank. It is also easy to see that T is (symmetric) positive definite on the null-space of N.

Consider now the following quadratic constrained minimization problem: Given $\mathbf{v} \in V_2$, find a $\mathbf{w} \in V_1$ that solves

$$\begin{array}{c} \frac{1}{2}\mathbf{w}^T T \mathbf{w} \mapsto \min \\ \text{subject to } N\mathbf{w} = \mathbf{v}. \end{array} \tag{2.2}$$

By forming the Lagrangian $\mathcal{L}(\mathbf{w},\ \lambda) = \frac{1}{2}\mathbf{w}^T T\mathbf{w} + \lambda^T(N\mathbf{w} - \mathbf{v})$ and setting its partial derivatives to zero we get the following saddle-point problem for $\mathbf{w}$ and the Lagrange multiplier λ,

$$\begin{bmatrix} T & N^T \\ N & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{v} \end{bmatrix}.$$

Our assumptions on T and N (namely, T being positive definite on the null-space of N and N^T having full column rank) tell us that the above problem has a unique solution $(\mathbf{w}_*,\ \lambda_*)$. It is also clear that the (negative) Schur complement $Z = N\left(T + N^T N\right)^{-1} N^T$ of the saddle-point matrix

$$\begin{bmatrix} T + N^T N & N^T \\ N & 0 \end{bmatrix},$$

is. s.p.d. (hence invertible). Since $(\mathbf{w}_*,\ \lambda_*)$ solves the equivalent problem,

$$\begin{bmatrix} T + N^T N & N^T \\ N & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} N^T \mathbf{v} \\ \mathbf{v} \end{bmatrix},$$

one gets the identity,

$$\mathbf{w}_* = \left(T + N^T N\right)^{-1} N^T(\mathbf{v} - \lambda_*),$$

which implies

$$\mathbf{v} = N\mathbf{w}_* = N\left(T + N^T N\right)^{-1} N^T(\mathbf{v} - \lambda_*) = Z(\mathbf{v} - \lambda_*).$$

Hence, $\mathbf{v} - \lambda_* = Z^{-1}\mathbf{v}$, and therefore $\mathbf{v}^T\mathbf{v} - \mathbf{v}^T\lambda_* = \mathbf{v}^T Z^{-1}\mathbf{v}$. The latter implies (using $\mathbf{v} = N\mathbf{w}_*$, and $N^T\lambda_* = -T\mathbf{w}_*$)

$$1 + \frac{\mathbf{w}_*^T T\mathbf{w}_*}{\mathbf{v}^T\mathbf{v}} = \frac{\mathbf{v}^T Z^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{v}}.$$

Next, since $\mathbf{w}_*$ solves the constrained minimization problem (2.2) we arrive at the desired identity (2.1).

□

3. Sharp spectral equivalence results

Our main goal in this section will be to obtain a suitable expression for the best constant K taking part in the spectral equivalence relations between A and B, for $B = B_{TL}$ and $B = B_{TG}$:

$$\mathbf{v}^T A\mathbf{v} \le \mathbf{v}^T B\mathbf{v} \le K\ \mathbf{v}^T A\mathbf{v}. \tag{3.1}$$

We will try to handle both cases (the hierarchical two-level and the two-grid one) simultaneously by introducing the notation: $\mathcal{M} \equiv J^T M J$ and $\mathcal{A} \equiv J^T A J$, where either $J = S$ or $J = I$. That is, either $\mathcal{M} = S^T M S = M_s$ and $\mathcal{A} = S^T A S = A_s$, or $\mathcal{M} = M$ and $\mathcal{A} = A$. With this, and considering also the case of $B = \widehat{B}^a_{TL}$, we can write the general error transfer operator for (1.2) as follows (the specific operators E_{TL} and E_{TG} were defined in Section 1.3):

$$I - B^{-1}A = E \equiv (I - J\mathcal{M}^{-T}J^T A)(I - P\mathcal{D}^{-1}P^T A)(I - J\mathcal{M}^{-1}J^T A), \tag{3.2}$$

where $\mathcal{D}$ is an s.p.d. approximation to the coarse-grid matrix $A_c = P^T AP$. Of particular importance is the case $\mathcal{D} = A_c$, which we consider in great detail, but we also point out how the convergence rate can be estimated using an appropriate approximation $\mathcal{D}$ to A_c.

Multiplying both sides of (3.2) by A we get that

$$AE = A^{\frac{1}{2}}(I - A^{\frac{1}{2}}J\mathcal{M}^{-T}J^T A^{\frac{1}{2}})(I - A^{\frac{1}{2}}P\mathcal{D}^{-1}P^T A^{\frac{1}{2}})(I - A^{\frac{1}{2}}J\mathcal{M}^{-1}J^T A^{\frac{1}{2}})A^{\frac{1}{2}}. \tag{3.3}$$

Let us denote for a moment $X = (I - A^{\frac{1}{2}}P\mathcal{D}^{-1}P^T A^{\frac{1}{2}})$. Note that if $\mathcal{D}$ is defined such that $\mathbf{v}_c^T\mathcal{D}\mathbf{v}_c \ge \mathbf{v}_c^T A_c\mathbf{v}_c$, then both X (see Lemma 3.1 for details) and AE are symmetric positive semi-definite. Equivalently, one has that $A^{\frac{1}{2}}EA^{-\frac{1}{2}} = I - A^{\frac{1}{2}}B^{-1}A^{\frac{1}{2}}$ is also symmetric positive semi–definite, which guarantees that the left-hand inequality in (3.1) holds assuming that B^{-1} is s.p.d. The latter follows from the fact that $\|E\|_A \le 1$, which is seen from (3.2) and (1.4).

Consider now the right-hand inequality in (3.1). Since AE is symmetric positive semi-definite, then $\|E\|_A$ is given by the largest eigenvalue of $A^{\frac{1}{2}}EA^{-\frac{1}{2}}$, or equivalently by $1 - \frac{1}{K}$. But, X being symmetric positive semi–definite implies that $X^{\frac{1}{2}}$ is well-defined, and it is obvious from a well–known fact (i.e., $\|G\| = \|G^T\|$, used for $G = X^{\frac{1}{2}}(I - A^{\frac{1}{2}}J\mathcal{M}^{-1}J^T A^{\frac{1}{2}})$) that the largest eigenvalue of

$$A^{\frac{1}{2}}EA^{-\frac{1}{2}} = (I - A^{\frac{1}{2}}J\mathcal{M}^{-T}J^T A^{\frac{1}{2}})X(I - A^{\frac{1}{2}}J\mathcal{M}^{-1}J^T A^{\frac{1}{2}})$$

is the same as the the largest eigenvalue of

$$\Theta \equiv X^{\frac{1}{2}}(I - A^{\frac{1}{2}}J\mathcal{M}^{-1}J^TA^{\frac{1}{2}})(I - A^{\frac{1}{2}}J\mathcal{M}^{-T}J^TA^{\frac{1}{2}})X^{\frac{1}{2}}.$$

Therefore, we will proceed with estimating the last expression.

We first consider the case $\mathcal{D} = A_c$. One notices then that $X = I - \overline{\pi}_A$ and the square root can be removed, because $X^2 = X$ (and hence $X = X^{\frac{1}{2}}$) in this case. Recalling the definition (1.5) of $\widetilde{M}$, consider then the expression

$$\begin{aligned}
\mathbf{v}^T\Theta\mathbf{v} &= (X\mathbf{v})^T\left(I - A^{\frac{1}{2}}J\left(\mathcal{M}^{-1} + \mathcal{M}^{-T} - \mathcal{M}^{-1}(J^TAJ)\mathcal{M}^{-T}\right)J^TA^{\frac{1}{2}}\right)(X\mathbf{v})\\
&= (X\mathbf{v})^T\left(I - A^{\frac{1}{2}}J\left(\mathcal{M}^{-1} + \mathcal{M}^{-T} - \mathcal{M}^{-1}\mathcal{A}\mathcal{M}^{-T}\right)\ J^TA^{\frac{1}{2}}\right)(X\mathbf{v})\\
&= (X\mathbf{v})^T\left(I - A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}\right)(X\mathbf{v})\\
&= \mathbf{v}^T(I - \overline{\pi}_A)^2\mathbf{v} - \mathbf{v}^T(I - \overline{\pi}_A)A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}(I - \overline{\pi}_A)\mathbf{v}\\
&= \mathbf{v}^T\left(I - \overline{\pi}_A - (I - \overline{\pi}_A)A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}(I - \overline{\pi}_A)\right)\mathbf{v}\\
&\leq \left(1 - \tfrac{1}{K}\right)\mathbf{v}^T\mathbf{v}.
\end{aligned}$$

The smallest (i.e. the best) constant K in the above inequality can then be defined via the following relation

$$\frac{1}{K} = \inf_{\mathbf{v}} \frac{\mathbf{v}^T\left(\overline{\pi}_A + (I - \overline{\pi}_A)A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}(I - \overline{\pi}_A)\right)\mathbf{v}}{\mathbf{v}^T\mathbf{v}}.$$

Since $\mathbf{v}^T\mathbf{v} = \mathbf{v}^T\overline{\pi}_A\mathbf{v} + \mathbf{v}^T(I - \overline{\pi}_A)\mathbf{v}$ with both components being non-negative, it is clear that the above minimum is taken over all $\mathbf{v} \in \text{Range}(I - \overline{\pi}_A)$. To see this we use the elementary inequality $f(t) \equiv \frac{t^2+b^2}{t^2+c^2} \geq f(0) = \frac{b^2}{c^2}$ if $b^2 \leq c^2$. Letting $t^2 = \mathbf{v}^T\overline{\pi}_A\mathbf{v}$, $c^2 = \mathbf{v}^T(I - \overline{\pi}_A)\mathbf{v}$ and $b^2 = \mathbf{v}^T(I - \overline{\pi}_A)A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}(I - \overline{\pi}_A)\mathbf{v}$, it remains to show that $b^2 \leq c^2$. The latter follows from the identity $\mathcal{A} = J^TAJ$ which implies that $\|G\| = \|G^T\| = 1$ for $G = \mathcal{A}^{-\frac{1}{2}}J^TA^{\frac{1}{2}}$. That is, with $\mathbf{w} = (I - \overline{\pi}_A)\mathbf{v}$, using the fact that $(\mathcal{A}^{-1} - \widetilde{\mathcal{M}}^{-1})$ is symmetric positive semi-definite (see (1.6)), one arrives at,

$$b^2 = \mathbf{w}^TA^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}\mathbf{w} \leq \mathbf{w}^TA^{\frac{1}{2}}J\mathcal{A}^{-1}J^TA^{\frac{1}{2}}\mathbf{w} = \mathbf{w}^TG^TG\mathbf{w} \leq \mathbf{w}^T\mathbf{w} = c^2.$$

Thus, we have

$$\frac{1}{K} = \inf_{\mathbf{v}} \frac{\mathbf{v}^T\left((I - \overline{\pi}_A)A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^TA^{\frac{1}{2}}(I - \overline{\pi}_A)\right)\mathbf{v}}{\mathbf{v}^T(I - \overline{\pi}_A)\mathbf{v}}. \tag{3.4}$$

Since above, in (3.4), we have $\widetilde{\mathcal{M}}^{-1}$, we will use Lemma 2.1 to get an estimate in terms of $\widetilde{\mathcal{M}}$ instead in the next theorem.

Theorem 3.1. *Assume that J and P are such that any vector $\mathbf{v}$ can be decomposed as $\mathbf{v} = J\mathbf{w} + P\mathbf{x}$. Then the best constant K in* (3.1) *is given by*

$$K = \sup_{\mathbf{v}\in Range(I-\pi_A)}\ \inf_{\mathbf{w}:\ \mathbf{v}=(I-\pi_A)J\mathbf{w}} \frac{\mathbf{w}^T\widetilde{\mathcal{M}}\mathbf{w}}{\mathbf{v}^TA\mathbf{v}}. \tag{3.5}$$

Proof. Let $N = (I - \overline{\pi}_A)A^{\frac{1}{2}}J$. Then

$$N^T N = J^T \left(A^{\frac{1}{2}}(I - \overline{\pi}_A)^2 A^{\frac{1}{2}}\right) J = J^T A(I - \pi_A)J.$$

Define

$$T = \widetilde{\mathcal{M}} - J^T A(I - \pi_A)J = \widetilde{\mathcal{M}} - \mathcal{A} + J^T A\pi_A J.$$

It is clear that T is symmetric positive semi-definite. We also have,

$$\widetilde{\mathcal{M}} = T + N^T N.$$

We point out here, that in order to apply Lemma 2.1, a minor (but very important) detail needs to be checked out; namely, whether or not the mapping $N : V_s \mapsto \text{Range}(I - \overline{\pi}_A)$, when $J = S$ is onto. This is done in the following way. Let $\mathbf{x} \neq 0$, $\mathbf{x} \in \text{Range}(I - \overline{\pi}_A)$ be given. We will find a $\mathbf{w} \in V_s$, such that $N\mathbf{w} = \mathbf{x}$. Since $\mathbf{x} \in \text{Range}(I - \overline{\pi}_A)$ there is a $\mathbf{v} \in V$, such that $\mathbf{x} = (I - \overline{\pi}_A)\mathbf{v}$. By the assumptions of the theorem, the following decompositions hold,

$$A^{-1/2}\mathbf{v} = J\mathbf{w} + P\mathbf{y}, \quad \text{and} \;\; \mathbf{v} = A^{1/2}J\mathbf{w} + A^{1/2}P\mathbf{y}.$$

Since $(I - \overline{\pi}_A)A^{\frac{1}{2}}P = 0$, one gets that $\mathbf{x} = N\mathbf{w}$. Then the "saddle-point" lemma 2.1 with $Z = N\left(T + N^T N\right)^{-1} N^T = (I - \overline{\pi}_A)A^{\frac{1}{2}}J\widetilde{\mathcal{M}}^{-1}J^T A^{\frac{1}{2}}(I - \overline{\pi}_A)$ and $V_2 = \text{Range}(I - \overline{\pi}_A)$, applied to identity (3.4), gives us the new identity

$$\frac{1}{K} = \inf_{\mathbf{v}\in V_2} \frac{\mathbf{v}^T Z\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \inf_{\mathbf{v}\in V_2} \frac{\mathbf{v}^T\mathbf{v}}{\mathbf{v}^T Z^{-1}\mathbf{v}} = \frac{1}{1 + \sup\limits_{\mathbf{v}\in V_2} \inf\limits_{\mathbf{w}:\, N\mathbf{w}=\mathbf{v}} \frac{\mathbf{w}^T T\mathbf{w}}{\mathbf{v}^T\mathbf{v}}}. \tag{3.6}$$

Further, in the denominator on the right hand side of (3.6) we have $\mathbf{v} = N\mathbf{w} = (I - \overline{\pi}_A)A^{\frac{1}{2}}J\mathbf{w}$. Replace now $\mathbf{v} \;\equiv A^{\frac{1}{2}}\mathbf{v}$. This implies that $N\mathbf{w} = A^{\frac{1}{2}}\mathbf{v}$, or $A^{\frac{1}{2}}\mathbf{v} = (I - \overline{\pi}_A)A^{\frac{1}{2}}J\mathbf{w}$; that is, $\mathbf{v} = (I - \pi_A)J\mathbf{w}$. Identity (3.6) then leads to (noticing that now $\mathbf{v} \in \text{Range}(I - \pi_A)$)

$$K = 1 + \sup_{\mathbf{v}\in \text{Range}(I - \pi_A)} \inf_{\mathbf{w}:\, \mathbf{v}=(I-\pi_A)J\mathbf{w}} \frac{\mathbf{w}^T T\mathbf{w}}{\mathbf{v}^T A\mathbf{v}}.$$

We also have,

$$\mathbf{w}^T T\mathbf{w} = \mathbf{w}^T \widetilde{\mathcal{M}}\mathbf{w} - \mathbf{w}^T N^T N\mathbf{w} = \mathbf{w}^T \widetilde{\mathcal{M}}\mathbf{w} - \mathbf{v}^T A\mathbf{v}.$$

Substituting the last expression in the above formula for K, implies the desired result (3.5). □

Corollary 3.1. *Assume that S and P provide a unique decomposition; namely, that $[S,\, P]$ is an invertible square matrix. Then,*

$$K = \sup_{\mathbf{w}} \frac{\mathbf{w}^T \widetilde{M}_s\mathbf{w}}{\mathbf{w}^T S^T(I - \pi_A)A(I - \pi_A)S\mathbf{w}}. \tag{3.7}$$

Proof. Note that for $\mathbf{v} = (I - \pi_A)S\mathbf{w}$, we also have $\mathbf{v} = S\mathbf{w} + P(-A_c^{-1}P^T A)(S\mathbf{w})$. Then, since S and P provide a unique decomposition of $\mathbf{v} = S\mathbf{w} + P\mathbf{x}$, this shows that the second component of $\mathbf{v}$ (which is in fact unique) equals $\mathbf{x} = -A_c^{-1}(P^T A)(S\mathbf{w})$. I.e., there is no inf in the formula for K. Thus,

$$K = \sup_{\mathbf{w}} \frac{\mathbf{w}^T \widetilde{M}_s\mathbf{w}}{\mathbf{w}^T S^T A(I - \pi_A)S\mathbf{w}},$$

which is the same as (3.7). □

3.1. Analysis of the two-level hierarchical basis preconditioner B_{TL}

We will now derive an upper bound for K in the case of S (i.e., $J = S$) and P providing a unique decomposition. We recall that $\mathcal{A} = S^T A S = A_s$. The following lemma gives the monotone dependence of $B - A$ on $\mathcal{D} - A_c$, and is needed in proving Theorem 3.2.

Lemma 3.1. *If $\mathcal{D} - A_c$ is symmetric positive semidefinite, then $B - A$ is symmetric positive semidefinite.*

Proof. Since $A_c = P^T A P$, and hence $\|G\| = \|G^T\| = 1$ for $G = A_c^{-\frac{1}{2}} P^T A^{\frac{1}{2}}$, we have that

$$\mathbf{v}^T A^{-1} \mathbf{v} \geq (P^T \mathbf{v})^T A_c^{-1} (P^T \mathbf{v}).$$

This and the assumptions of the lemma lead to the inequality

$$\mathbf{v}^T A^{-1} \mathbf{v} \geq (P^T \mathbf{v})^T \mathcal{D}^{-1} (P^T \mathbf{v}),$$

which is equivalent to

$$\mathbf{v}^T \left(I - A^{\frac{1}{2}} P \mathcal{D}^{-1} P^T A^{\frac{1}{2}} \right) \mathbf{v} \geq 0.$$

Hence, $AE = A(A^{-1} - B^{-1})A$ is symmetric positive semi-definite, which is equivalent to $B - A$ being symmetric positive semi-definite. $\square$

Theorem 3.2. *Assume that M_s provides a convergent splitting for A_s in the A_s-inner product, i.e., that $\left(M_s + M_s^T - A_s\right)$ is s.p.d. Also, let $\gamma \in [0, 1)$ be the constant in the strengthened Cauchy-Schwarz inequality (1.8). Then B_{TL} (i.e., B with $\mathcal{D} = A_c$; see Definition 1.1) and A are spectrally equivalent and the following bounds hold:*

$$\mathbf{v}^T A \mathbf{v} \leq \mathbf{v}^T B_{TL} \mathbf{v} \leq K \mathbf{v}^T A \mathbf{v}, \quad K \leq \frac{1}{1 - \gamma^2} \sup_{\mathbf{w}} \frac{\mathbf{w}^T \widetilde{M}_s \mathbf{w}}{\mathbf{w}^T A_s \mathbf{w}}. \tag{3.8}$$

In the case of B with an inexact second block $\mathcal{D}$ that satisfies

$$0 \leq \mathbf{x}^T (\mathcal{D} - A_c) \mathbf{x} \leq \delta \, \mathbf{x}^T A_c \mathbf{x},$$

the following perturbation result holds,

$$\mathbf{v}^T A \mathbf{v} \leq \mathbf{v}^T B \mathbf{v} \leq \left(K + \frac{\delta}{1 - \gamma^2} \right) \mathbf{v}^T A \mathbf{v}.$$

Proof. The estimate (3.8) follows from (3.7) and (1.9), which combined give the following upper bound for K,

$$K \leq \sup_{\mathbf{w}} \frac{\mathbf{w}^T \widetilde{M}_s \mathbf{w}}{\mathbf{w}^T A_s \mathbf{w}} \, \frac{1}{1 - \gamma^2},$$

The proof is completed by using Lemma 3.1, (1.11), and some obvious inequalities as follows:

$$\begin{aligned} 0 \leq \mathbf{v}^T (B - A) \mathbf{v} \quad &= \mathbf{v}^T (B_{TL} - A) \mathbf{v} + \mathbf{x}^T (\mathcal{D} - A_c) \mathbf{x}, \quad \mathbf{x} = P^T \mathbf{v}, \\ &\leq \mathbf{v}^T (B_{TL} - A) \mathbf{v} + \delta \mathbf{x}^T A_c \mathbf{x} \\ &\leq \mathbf{v}^T (B_{TL} - A) \mathbf{v} + \tfrac{\delta}{1 - \gamma^2} \, \mathbf{x}^T \mathcal{S}_A \mathbf{x} \\ &\leq \mathbf{v}^T (B_{TL} - A) \mathbf{v} + \tfrac{\delta}{1 - \gamma^2} \, \mathbf{v}^T A \mathbf{v} \\ &\leq \left(K - 1 + \tfrac{\delta}{1 - \gamma^2} \right) \mathbf{v}^T A \mathbf{v}. \end{aligned}$$

□

The latter result is a generalization of a main two-level convergence theorem found in [2]. A more detailed study of the (two-level) approximate block-factorization preconditioners B^a_{TL} (defined as in (1.12)) is found in Chapter 9 of [1]. Note that in the classical two-level HB methods, A_s happens to be well-conditioned, so it is not impractical to let $M_s = A_s$. Then, $K = \frac{1}{1-\gamma^2}$ (if $\mathcal{D} = A_c$), i.e., the two-level convergence factor $\varrho(E_{TL}) = 1 - \frac{1}{K} = \gamma^2$ equals to the $\cos^2$ of the abstract angle between the hierarchical components Range(S) and Range(P), a well-known classical result.

3.2. Analysis of the two-grid preconditioner B_{TG}

Here, we consider the case of $B = B_{TG}$, i.e., $J = I$. The analysis follows the lines of the analysis from the previous subsection 3.1, and in a similar fashion one obtains that the smallest constant K is given by the identity (see (3.5) in Theorem 3.1)

$$K = \sup_{\mathbf{v} \in \mathrm{Range}(I-\pi_A)} \ \inf_{\mathbf{w}:\ \mathbf{v}=(I-\pi_A)\mathbf{w}} \frac{\mathbf{w}^T \widetilde{M} \mathbf{w}}{\mathbf{v}^T A \mathbf{v}}.$$

That is,

$$K = \sup_{\mathbf{v}} \frac{\inf_{\mathbf{w}} \left(\pi_A \mathbf{w} + (I-\pi_A)\mathbf{v}\right)^T \widetilde{M} \left(\pi_A \mathbf{w} + (I-\pi_A)\mathbf{v}\right)}{\left((I-\pi_A)\mathbf{v}\right)^T A \left((I-\pi_A)\mathbf{v}\right)}.$$

The inf over $\mathbf{w}$ is attained at $\mathbf{w}:\ \pi_A(\mathbf{v}-\mathbf{w}) = \pi_{\widetilde{M}}\mathbf{v}$, that is,

$$A_c^{-1} P^T A (\mathbf{v}-\mathbf{w}) = \widetilde{M}_c^{-1} P^T \widetilde{M} \mathbf{v}.$$

Here $\widetilde{M}_c = P^T \widetilde{M} P$ and $\pi_{\widetilde{M}} = P \widetilde{M}_c^{-1} P^T \widetilde{M}$. For a $\mathbf{w} = P\mathbf{w}_c$ one has $A_c^{-1} P^T A \mathbf{w} = \mathbf{w}_c$. Hence,

$$\mathbf{w}_c = A_c^{-1} P^T A \mathbf{v} - \widetilde{M}_c^{-1} P^T \widetilde{M}\ \mathbf{v}.$$

This implies, $\pi_A \mathbf{w} = P\mathbf{w}_c = (\pi_A - \pi_{\widetilde{M}})\mathbf{v}$. Therefore, $\pi_A \mathbf{w} + (I-\pi_A)\mathbf{v} = (I - \pi_{\widetilde{M}})\mathbf{v}$. Note also that the following relation holds $(I-\pi_{\widetilde{M}})(I-\pi_A) = I - \pi_{\widetilde{M}}$ and $\mathbf{v}^T A (I-\pi_A)\mathbf{v} \le \mathbf{v}^T A \mathbf{v}$. Thus we arrive at the final estimate which is formulated in the next theorem.

Theorem 3.3. *The convergence factor of the two-grid method, $\varrho(E_{TG}) = 1 - \frac{1}{K}$, is characterized by*

$$K = \sup_{\mathbf{v}} \frac{\left((I-\pi_{\widetilde{M}})\mathbf{v}\right)^T \widetilde{M} (I-\pi_{\widetilde{M}})\mathbf{v}}{((I-\pi_A)\mathbf{v})^T A (I-\pi_A)\mathbf{v}} = \sup_{\mathbf{v}} \frac{\mathbf{v}^T \widetilde{M} (I-\pi_{\widetilde{M}})\mathbf{v}}{\mathbf{v}^T A \mathbf{v}}. \tag{3.9}$$

Remark 3.1. *Consider the case when the coarse degrees of freedom are defined on the basis of a mapping $R:\ \mathbb{R}^n \mapsto \mathbb{R}^{n_c}$ such that $RP = I$. Let $Q = PR$. Note that Q is a projection (i.e., $Q^2 = Q$). Next, noticing that*

$$\|(I-\pi_{\widetilde{M}})\mathbf{e}\|^2_{\widetilde{M}} = \inf_{\mathbf{v} \in Range(P)} \|\mathbf{e}-\mathbf{v}\|^2_{\widetilde{M}} \le \|\mathbf{e} - PR\mathbf{e}\|^2_{\widetilde{M}},$$

one gets the following upper bound $K \le \sup_{\mathbf{e}} \ \mu_{\widetilde{M}}(Q,\ \mathbf{e})$, where the latter quantity (sometimes called measure) equals,

$$\mu_{\widetilde{M}}(Q,\ \mathbf{e}) = \frac{\langle \widetilde{M}(I-Q)\mathbf{e},\ (I-Q)\mathbf{e}\rangle}{\langle A\mathbf{e}, \mathbf{e}\rangle}.$$

The latter measure $\mu_{\widetilde{M}}(Q,\ \mathbf{e})$ was the main ingredient used in [11] to find sufficient conditions for the two-grid convergence of the respective AMG. With the above Theorem 3.3 in hand, we are now able to show that the conditions from [11] are also necessary (in the sense of the next two corollaries 3.2–3.3).

Corollary 3.2. *For any $\mathbf{v}$ in the space Range$(I - \pi_{\widetilde{M}})$ one has the spectral equivalence relations,*

$$\mathbf{v}^T A\mathbf{v} \leq \mathbf{v}^T \widetilde{M}\mathbf{v} \leq K\ \mathbf{v}^T A\mathbf{v}.$$

That is, in a space complementary to the coarse space Range(P) the symmetrized smoother $\widetilde{M}$ is an efficient preconditioner for A. If one introduces the matrix S such that Range$(S) =$ Range$(I - \pi_{\widetilde{M}})$ one has the following spectral equivalence relations between $A_s = S^T AS$ and $\widetilde{M}_s = S^T \widetilde{M} S$,

$$\mathbf{v}_s^T A_s \mathbf{v}_s \leq \mathbf{v}_s^T \widetilde{M}_s \mathbf{v}_s \leq K\ \mathbf{v}_s^T A_s \mathbf{v}_s.$$

Proof. Using the inequalities

$$\|(I - \pi_A)\mathbf{w}\|_A = \inf_{\mathbf{v} \in \text{Range}(P)} \|\mathbf{w} - \mathbf{v}\|_A \leq \|(I - \pi_{\widetilde{M}})\mathbf{w}\|_A,$$

and $\mathbf{v}^T A\mathbf{v} \leq \mathbf{v}^T \widetilde{M}\mathbf{v}$, the result is straightforward. □

Letting $R_* = \widetilde{M}_c^{-1} P^T \widetilde{M}$ one candidate for an S that spans the space Range$(I - \pi_{\widetilde{M}})$ is $S = I - PR_*$. One notices that $R_* P = I$ (and $R_* S = 0$), i.e., we are in the setting of Remark 3.1 with $R = R_*$. We point out that R_* is the optimal mapping that defines the coarse degrees of freedom, in the sense that such a choice will provide the best convergence rate. Unfortunately, constructing R_* in general may involve work comparable to the work needed to solve the original problem (1.1). This is due to the fact that, for given sparse operators P and M (or its symmetrized version $\widetilde{M}$), the optimal R is not in general sparse, due to presence of $\widetilde{M}_c^{-1}$ in its definition. However, in the following important example, R_* happens to be the simple injection mapping.

Example 3.1. *Consider a two–by–two block partitioning of A,*

$$A = \begin{bmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix},$$

corresponding to a "f" and "c" block–partitioning of the vectors $\mathbf{v} = \begin{bmatrix} \mathbf{v}_f \\ \mathbf{v}_c \end{bmatrix}$ in V. Introduce the splitting $A = D_A - L - U$, with

$$D_A = \begin{bmatrix} A_{ff} & 0 \\ 0 & A_{cc} \end{bmatrix},\ -L = \begin{bmatrix} 0 & 0 \\ A_{cf} & 0 \end{bmatrix},\ \textit{and } U = L^T.$$

For two given (not necessarily symmetric) matrices M_{ff} and M_{cc}, consider the interpolation matrix

$$P = \begin{bmatrix} -M_{ff}^{-1} A_{fc} \\ I \end{bmatrix},$$

and the (inexact) block Gauss–Seidel smoother (also called "c"–"f" relaxation)

$$M = \begin{bmatrix} M_{ff} & A_{fc} \\ 0 & M_{cc} \end{bmatrix} = D - U,\ D = \begin{bmatrix} M_{ff} & 0 \\ 0 & M_{cc} \end{bmatrix}$$

Then, $R_* = \widetilde{M}_c^{-1} P^T \widetilde{M} = [0,\ I]$, *i.e.,* R_* *is the trivial injection mapping. The optimal* S *defined as* $S_* = I - \pi_{\widetilde{M}}$ *takes the simple form* $S_* = I - PR_* = \left[\begin{array}{cc} I & M_{ff}^{-1}A_{fc} \\ 0 & 0 \end{array}\right]$. *The latter shows that* $Range(S_*) = Range \left[\begin{array}{c} I \\ 0 \end{array}\right]$. *Finally, the corresponding two–grid convergence factor* $\varrho(E_{TG}) = 1 - \frac{1}{K}$ *is characterized with the identity (assuming that* $\|I - M^{-1}A\|_A < 1$*),*

$$K = \sup_{\mathbf{e}}\ \mu_{\widetilde{M}}(Q,\ \mathbf{e}),$$

where $Q = PR_*$*, and* $\mu_{\widetilde{M}}(Q,\ \mathbf{e}) = \frac{\langle \widetilde{M}(I-Q)\mathbf{e},\ (I-Q)\mathbf{e}\rangle}{\langle A\mathbf{e},\ \mathbf{e}\rangle}$ *is the measure used in [11].*

Proof. One has, assuming that all inverses below exist,

$$\begin{aligned}
P^T\widetilde{M} &= \left[-A_{cf}M_{ff}^{-T},\ I\right] M^T\left(M + M^T - A\right)^{-1} M \\
&= \left[-A_{cf}M_{ff}^{-T},\ I\right] \left[\begin{array}{cc} M_{ff}^T & 0 \\ A_{cf} & M_{cc}^T \end{array}\right] \left(D^T + D - D_A\right)^{-1} M \\
&= \left[0,\ M_{cc}^T\right] \left(D^T + D - D_A\right)^{-1} M \\
&= \left[0,\ M_{cc}^T\left(M_{cc}^T + M_{cc} - A_{cc}\right)^{-1}\right] \left[\begin{array}{cc} M_{ff} & A_{fc} \\ 0 & M_{cc} \end{array}\right] \\
&= \left[0,\ M_{cc}^T\left(M_{cc}^T + M_{cc} - A_{cc}\right)^{-1} M_{cc}\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\widetilde{M}_c &= P^T\widetilde{M}P = \left[0,\ M_{cc}^T\left(M_{cc}^T + M_{cc} - A_{cc}\right)^{-1} M_{cc}\right] \left[\begin{array}{c} -M_{ff}^{-1}A_{fc} \\ I \end{array}\right] \\
&= M_{cc}^T\left(M_{cc}^T + M_{cc} - A_{cc}\right)^{-1} M_{cc}.
\end{aligned}$$

Finally,

$$\begin{aligned}
R_* &= \widetilde{M}_c^{-1} P^T \widetilde{M} \\
&= M_{cc}^{-1}(M_{cc}^T + M_{cc} - A_{cc})M_{cc}^{-T} \left[0,\ M_{cc}^T\left(M_{cc}^T + M_{cc} - A_{cc}\right)^{-1} M_{cc}\right] \\
&= [0,\ I].
\end{aligned}$$

That is, $R = R_*$ is the simple injection mapping. The rest follows from the definition of $S_* = I - \pi_{\widetilde{M}} = I - PR_* = I - Q$ and Theorem 3.3. □

In the general case, for any well-conditioned smoother $\widetilde{M}$ (as the standard ones are), the entries of $\widetilde{M}_c^{-1}$ will have a fast decay rate away from the diagonal (similar to the inverse of a finite element mass-matrix). Therefore, reasonably sparse approximations R to R_* will be available, and hence using such a sparse R in practice will be a feasible and accurate enough choice. Of course, using an approximation, as indicated in Remark 3.1, will only give upper bounds for K. More specifically, we will have $K \leq \frac{1}{\lambda_{\min}}$ by finding (accurate) bounds of the minimal eigenvalue $\lambda_{\min}$ of the generalized eigenvalue problem

$$A\mathbf{q} = \lambda(I - Q^T)\widetilde{M}(I - Q)\mathbf{q}$$

where $Q = PR$. Notice that the above eigenvalue problem takes a particularly appealing form in the case of Example 3.1, but this is left for future study and is not considered further in this paper. A somewhat simpler approach for estimating K is found in Theorem 4.1.

Corollary 3.3. *The following three statements are equivalent, and they are necessary and sufficient conditions for uniform convergence of the two grid method:*

1. $(I - \pi_{\widetilde{M}})$ *is bounded in A-norm*

$$\left((I - \pi_{\widetilde{M}})\mathbf{v}\right)^T A(I - \pi_{\widetilde{M}})\mathbf{v} \le K\ \mathbf{v}^T A\mathbf{v}.$$

2. $\pi_{\widetilde{M}}$ *is bounded in A-norm*

$$\left(\pi_{\widetilde{M}}\mathbf{v}\right)^T A\pi_{\widetilde{M}}\mathbf{v} \le K\ \mathbf{v}^T A\mathbf{v}.$$

3. *The spaces Range*$(S) \equiv$ *Range*$(I - \pi_{\widetilde{M}})$ *and Range*(P) *have a non-trivial angle in A-inner product, that is,*

$$\left(\mathbf{v}_s^T S^T AP\mathbf{x}\right)^2 \le (1 - \frac{1}{K})\ \mathbf{v}_s^T A_s\mathbf{v}_s\ \mathbf{x}^T A_c\mathbf{x},\ \textit{for any}\ \mathbf{v}_s\ \textit{and}\ \mathbf{x}.$$

Proof. We first give an argument that the three statements are equivalent. Consider the quadratic form $Q(t) = (\pi_{\widetilde{M}}\mathbf{v} + tS\mathbf{v}_s)^T A(\pi_{\widetilde{M}}\mathbf{v} + tS\mathbf{v}_s) - \frac{1}{K}\left(\pi_{\widetilde{M}}\mathbf{v}\right)^T A\pi_{\widetilde{M}}\mathbf{v}$. Note, that $\pi_{\widetilde{M}}S\mathbf{v}_s = 0$, hence $\pi_{\widetilde{\mathcal{M}}}(\mathbf{v} + tS\mathbf{v}_s) = \pi_{\widetilde{M}}\mathbf{v}$. This shows that $Q(t) \ge 0$ for any real t if $\pi_{\widetilde{M}}$ is A-bounded. Then the fact that its discriminant is non-positive, shows the strengthened Cauchy-Schwarz inequality since Range$(\pi_{\widetilde{M}})$ = Range(P). The argument goes both ways. Namely, the strengthened Cauchy-Schwarz inequality implies that the discriminant is non-positive, hence Q is non-negative, that is, $\pi_{\widetilde{M}}$ is bounded in energy. Due to the symmetry of the strengthened Cauchy-Schwarz inequality one sees that $I - \pi_{\widetilde{M}}$ has the same energy norm as $\pi_{\widetilde{M}}$ if $K > 1$. That these are necessary and sufficient follows from Theorem 3.3. □

4. Algebraic two-grid methods and preconditioners

Corollaries 3.2 and 3.3 represent the main foundation for constructing efficient two-grid preconditioners. Namely, one needs a coarse space Range(P) such that there is a complementary one, Range(S), with the properties:

(i) the symmetrized smoother restricted to the subspace Range(S), i.e., $S^T\widetilde{M}S$, is spectrally equivalent to the subspace matrix S^TAS, and
(ii) the complementary spaces Range(S) and Range(P) have a non-trivial angle in A-inner product; that is, they are almost A-orthogonal.

In practice, one needs a sparse matrix P so that the coarse matrix P^TAP is also sparse, whereas explicit knowledge of the best S is not really needed. If P and S are constructed based solely on A, and similarly the smoother M (or the symmetrized one, $\widetilde{M}$) comes from a convergent splitting of A, then the resulting method (or preconditioner) belongs to the class of "algebraic" two-grid methods (or preconditioners) or simply AMG (when several coarsening levels are used). For some basic facts about AMG we refer the reader to, e.g., Ruge and Stüben [12], or the tutorial [8].

In order to guarantee the efficiency of the method, one only needs an S (not necessarily the best one defined as Range$(I - \pi_{\widetilde{M}})$) in order to test if the subspace smoother $S^T\widetilde{M}S$ is efficient on the subspace matrix S^TAS. That is, one needs an estimate (for the particular S)

$$\mathbf{v}_s^T A_s\mathbf{v}_s \le \mathbf{v}_s^T\widetilde{M}_s\mathbf{v}_s \le \kappa\ \mathbf{v}_s^T A_s\mathbf{v}_s, \tag{4.1}$$

with a reasonable constant κ. The efficiency of the smoother on the complementary space Range(S) is sometimes referred to as efficient *compatible relaxation*. The latter notion is due to Achi Brandt [7].

The second main ingredient is the energy boundedness of P in the sense that for a small constant η one wants the bound,

$$\mathbf{x}^T A_c \mathbf{x} \leq \eta \inf_{\mathbf{v}_s:\ \mathbf{v}=S\mathbf{v}_s+P\mathbf{x}} \mathbf{v}^T A \mathbf{v}. \tag{4.2}$$

For the simple example of orthogonal $R^T = [\ 0 \ \ I\ ]^T$ and $S = [\ I \ \ 0\ ]^T$, and $P:\ RP = I$, one can show (see [11]) that (4.2) is equivalent to $Q \equiv PR$ being bounded in energy, i.e.,

$$\mathbf{v}^T Q^T A Q \mathbf{v} \leq \eta\ \mathbf{v}^T A \mathbf{v}.$$

Since $Q^2 = Q$, i.e., Q is a projection, the above estimate is equivalent to

$$((I-Q)\mathbf{v})^T A (I-Q)\mathbf{v} \leq \eta\ \mathbf{v}^T A \mathbf{v}.$$

It is clear then, that a sufficient condition for P to be bounded is to establish the following "weak approximation property",

$$\|A\| \|\mathbf{v} - PR\mathbf{v}\|^2 \leq \eta\ \mathbf{v}^T A \mathbf{v},$$

which was a common tool used in the classical two-(and multi-)grid convergence theory. One can actually prove the following main result in the general case (see Theorem 4.2 in [11]).

Theorem 4.1. *Assume that the estimates* (4.1) *and* (4.2) *hold true. Then the two-grid preconditioner* $B = B_{TG}$ *is spectrally equivalent to* A *with a constant* $K \leq \eta\kappa$.

Proof. We have to estimate K defined in (3.9). Since Range(S) is complementary to Range(P) (by assumption), then any $\mathbf{v}$ can be uniquely decomposed as $\mathbf{v} = S\mathbf{v}_s + P\mathbf{x}$. The term in the numerator of (3.9) can be estimated as follows,

$$\begin{aligned} ((I-\pi_{\widetilde{M}})\mathbf{v})^T \widetilde{M} ((I-\pi_{\widetilde{M}})\mathbf{v} &= \inf_{\mathbf{y}}\ (\mathbf{v} - P\mathbf{y})^T \widetilde{M} (\mathbf{v} - P\mathbf{y}) \\ &\leq (\mathbf{v} - P\mathbf{x})^T \widetilde{M} (\mathbf{v} - P\mathbf{x}) \\ &= \mathbf{v}_s^T S^T \widetilde{M} S \mathbf{v}_s \\ &\leq \kappa\ \mathbf{v}_s^T S^T A S \mathbf{v}_s. \end{aligned}$$

In the last line above we used (4.1).

The energy boundedness of (4.2) implies a strengthened Cauchy-Schwarz inequality for Range(S) and Range(P). That inequality is equivalent to the following energy boundedness of S,

$$\mathbf{v}_s^T S^T A S \mathbf{v}_s \leq \eta \inf_{\mathbf{x}:\ \mathbf{v}=S\mathbf{v}_s+P\mathbf{x}} \mathbf{v}^T A \mathbf{v}.$$

Using the projection π_A, one gets

$$\mathbf{v}_s^T S^T A S \mathbf{v}_s \leq \eta\ ((I-\pi_A)S\mathbf{v}_s)^T A (I-\pi_A) S \mathbf{v}_s.$$

Finally, since $(I-\pi_A)P\mathbf{x} = 0$, one arrives at the following bound for the denominator of (3.9),

$$\begin{aligned} \mathbf{v}_s^T S^T A S \mathbf{v}_s &\leq \eta\ ((I-\pi_A)(S\mathbf{v}_s + P\mathbf{x}))^T A (I-\pi_A)(S\mathbf{v}_s + P\mathbf{x}) \\ &= \eta\ ((I-\pi_A)\mathbf{v})^T A (I-\pi_A)\mathbf{v}. \end{aligned}$$

Thus, (3.9) is finally estimated as follows

$$K = \sup_{\mathbf{v}} \frac{((I-\pi_{\widetilde{M}})\mathbf{v})^T \widetilde{M}(I-\pi_{\widetilde{M}})\mathbf{v}}{((I-\pi_A)\mathbf{v})^T A(I-\pi_A)\mathbf{v}} \leq \sup_{\mathbf{v}_s} \frac{\kappa\, \mathbf{v}_s^T S^T A S \mathbf{v}_s}{\frac{1}{\eta}\mathbf{v}_s^T S^T A S \mathbf{v}_s} = \kappa\, \eta.$$

□

5. Window based spectral AMG

In the present section we provide a purely algebraic way of selecting coarse degrees of freedom and a way to construct an energy bounded interpolation matrix P. In the analysis, we will use a simple Richardson iteration as a smoother. The presented method is an "element-free" version of the spectral AMGe method studied in Chartier et al. [9]. All definitions and constructions below are valid in the case when A is positive and only semidefinite, i.e. may have nonempty null space $\mathrm{Null}(A)$.

We consider the problem (1.1) and reformulate it in the following equivalent least squares minimization:

$$\mathbf{u} = \arg\min_{\mathbf{v}} \sum_w \|A_w\mathbf{v} - \mathbf{f}_w\|^2_{D_w}. \tag{5.1}$$

In the least squares formulation, each w is a subset of $\{1,\dots,n\}$, and we assume that

$$\cup w = \{1,\dots,n\},$$

where the decomposition can be overlapping. The sets w are called windows, and represent a grouping of the rows of A. The corresponding rectangular matrices we denote by A_w, i.e., $A_w = \{A_{ij}\}_{i\in w, j=1:n}$. Thus we have that $A_w \in \mathbb{R}^{|w|\times n}$, where $|\cdot|$ stands for cardinality. Accordingly in (5.1) $\mathbf{f}_w = \mathbf{f}|_w = \{\mathbf{f}_i\}_{i\in w}$, denotes a restriction of $\mathbf{f}$ to a subset and $D_w = (D_w(i))_{i\in w}$ are diagonal matrices such that for any i, $\sum_{w:\, i\in w} D_w(i) = 1$, that is, $\{D_w\}_w$ provide a partition of unity. Vanishing the first variation of the least squares functional, we obtain that the solution to the minimization problem (5.1) satisfies

$$\sum_w (A_w)^T D_w A_w \mathbf{u} = \sum_w (A_w)^T D_w \mathbf{f}_w. \tag{5.2}$$

We will first estimate the maximum eigenvalue of the matrix on left hand side of (5.2) in terms of $\|A\|$. Let $A^T = [\mathbf{b}_1,\ \mathbf{b}_2,\ \dots,\ \mathbf{b}_n]$, where $\mathbf{b}_i^T$ is the ith row of A. The following identities hold:

$$\begin{aligned}
\mathbf{v}^T A^T A \mathbf{v} &= \sum_i (\mathbf{b}_i^T\mathbf{v})^2, \quad \text{and}\\
\sum_w \mathbf{v}^T (A_w)^T D_w A_w \mathbf{v} &= \sum_w \sum_{i\in w} D_w(i)(\mathbf{b}_i^T\mathbf{v})^2\\
&= \sum_i (\mathbf{b}_i^T\mathbf{v})^2 \left(\sum_{w:\, i\in w} D_w(i)\right)\\
&= \sum_i (\mathbf{b}_i^T\mathbf{v})^2.
\end{aligned}$$

Then we obtain that

$$\mathbf{v}^T\left(\sum_w (A_w)^T D_w A_w\right)\mathbf{v} = \mathbf{v}^T A^T A \mathbf{v}. \tag{5.3}$$

We emphasize that we will not solve the equivalent least squares problem (5.2) and it has only been introduced as a motivation to consider the "local" matrices $(A_w)^T D_w A_w$ as a tool for constructing sparse (and hence local) interpolation mapping P which we explain below. Of interest will be the Schur complements S_w, that are obtained from the matrices $(A_w)^T D_w A_w$ by eliminating the entries outside w. Note that S_w is symmetric and positive semidefinite and hence one has the inequality

$$(\mathbf{v}_w)^T S_w \mathbf{v}_w \leq \mathbf{v}^T (A_w)^T D_w A_w \mathbf{v}, \quad \mathbf{v}_w = \mathbf{v}|_w \,,$$

hence

$$\sum_w (\mathbf{v}_w)^T S_w \mathbf{v}_w \leq \mathbf{v}^T \left(\sum_w (A_w)^T D_w A_w \right) \mathbf{v} = \ \mathbf{v}^T A^T A \mathbf{v} \leq \ \|A\| \ \mathbf{v}^T A \mathbf{v}. \tag{5.4}$$

This inequality implies (letting $\mathbf{v} = 0$ outside w) that $(\mathbf{v}_w)^T S_w \mathbf{v}_w \leq \|A\| \ \mathbf{v}^T A \mathbf{v} \leq \|A\|^2 \ \mathbf{v}^T \mathbf{v} = \|A\|^2 \ \mathbf{v}_w^T \mathbf{v}_w$, that is,

$$\|S_w\| \leq \|A\|^2. \tag{5.5}$$

Selecting coarse degrees of freedom

Our goal will be to select a coarse space. The way we do that will be by fixing a window and associating with it a number $m_w \leq |w|$. Then we construct m_w basis vectors (columns of P) corresponding to this window in the following way: All the eigenvectors and eigenvalues of S_w are computed and the eigenvectors corresponding to the first m_w eigenvalues are chosen. Since generally the windows have overlap, another partition of unity is constructed, with diagonal matrices $\{Q_w\}$ where each Q_w is non-zero only on w and the set $\{Q_w\}$ satisfies $\sum\limits_w Q_w = I$. From the first m_w eigenvectors of S_w extended by zero outside w we form column-wise the local interpolation matrix P_w which hence has m_w columns. The global interpolation matrix is then defined as

$$P = \sum_w Q_w \, [0, \ P_w, \ 0] \,.$$

Here, for a global coarse vector $\mathbf{v}^c = (\mathbf{v}^c_w)$, the action of $[0, \ P_w, \ 0]$ is defined such that $[0, \ P_w, \ 0] \, \mathbf{v}^c = P_w (\mathbf{v}^c|_w) = P_w \mathbf{v}^c_w$.

The remainder of this section follows the presentation in [9] but the main result (Theorem 5.1) utilizes our main identity (3.9). The first result concerns the null-space of A. Namely, that it is contained in the range of the interpolation P.

Lemma 5.1. *Suppose that m_w is such that $m_w \geq$ dim Null (S_w) for every window w. Then Null $(A) \subset$ Range (P), that is, if $A\mathbf{v} = 0$, then there exists a $\mathbf{v}^c \in \mathbb{R}^{n_c}$ such that $\mathbf{v} = P\mathbf{v}^c$.*

Proof. Let $A\mathbf{v} = 0$. Then from inequality (5.4) it follows that $S_w \mathbf{v}_w = 0$, where $\mathbf{v}_w = \mathbf{v}|_w$ and we extend $\mathbf{v}_w$ by zero outside w whenever needed. Hence, by our assumption on m_w there exists a local coarse grid vector $\mathbf{v}^c_w$ such that $\mathbf{v}_w = P_w \mathbf{v}^c_w$. Let $\mathbf{v}^c$ be the composite coarse grid vector that agrees with $\mathbf{v}^c_w$ on w, for each w. This is simply the collection $\mathbf{v}^c = (\mathbf{v}^c_w)_s$. Then,

$$P\mathbf{v}^c = \sum_w Q_w P_w \mathbf{v}^c_w = \sum_w Q_w \mathbf{v}_w = \sum_w Q_w \mathbf{v} = \mathbf{v}.$$

$\square$

Two-grid convergence

First, we prove a main coarse-grid "weak approximation property".

Lemma 5.2. *Assume that the windows $\{w\}$ are selected in a "quasi-uniform" manner such that for all w the following uniform estimate holds*

$$\|S_w\| \geq \eta \|A\|^2.$$

Note that $\eta \leq 1$ (see (5.5)). Assume that we have chosen m_w so well that for a constant $\delta > 0$ uniformly in w one has

$$\|S_w\| \leq \delta \lambda_{m_w+1}(S_w),$$

where $\lambda_{m_w+1}(S_w)$ denotes the (m_w+1)th smallest eigenvalue of S_w. It is clear that $\delta \geq 1$. Then, for any vector $\mathbf{e} \in \mathbb{R}^n$, there exists a global interpolant ϵ in the range of P such that

$$(\mathbf{e}-\epsilon)^T A(\mathbf{e}-\epsilon) \leq \|A\| \|\mathbf{e}-\epsilon\|^2 \leq \frac{\delta}{\eta}\, \mathbf{e}^T A\mathbf{e}. \tag{5.6}$$

Proof. The analysis follows [9]. Let $\mathbf{e} \in \mathbb{R}^n$ be given. Note that our assumption on m_w is equivalent to the assumption that, for any window w, there exists a ϵ_w in the range of P_w such that

$$\|S_w\| \|\mathbf{e}_w - \epsilon_w\|^2 \leq \delta\, \mathbf{e}_w^T S_w \mathbf{e}_w, \tag{5.7}$$

where $\mathbf{e}_w = \mathbf{e}|_w$ and whenever needed we consider $\mathbf{e}_w$ and ϵ_w extended by zero outside w. We now construct an ϵ in the range of P which will satisfy (5.6). Namely, we set $\epsilon = \sum_w Q_w \epsilon_w$. One notices that $\sum_w Q_w \epsilon = \epsilon = \sum_w Q_w \epsilon_w$. Hence,

$$\begin{aligned}
\|\mathbf{e}-\epsilon\|^2 &= (\mathbf{e}-\epsilon)^T \left(\sum_w Q_w(\mathbf{e}-\epsilon)\right) \\
&= (\mathbf{e}-\epsilon)^T \left(\sum_w Q_w(\mathbf{e}_w-\epsilon_w)\right) \\
&= \sum_w \left(Q_w^{\frac{1}{2}}(\mathbf{e}-\epsilon)\right)^T \left(Q_w^{\frac{1}{2}}(\mathbf{e}_w-\epsilon_w)\right) \\
&\leq \left[\sum_w (\mathbf{e}-\epsilon)^T Q_w (\mathbf{e}-\epsilon)\right]^{\frac{1}{2}} \left[\sum_w \|Q_w^{\frac{1}{2}}(\mathbf{e}_w-\epsilon_w)\|^2\right]^{\frac{1}{2}} \\
&= \|\mathbf{e}-\epsilon\| \left[\sum_w \|Q_w^{\frac{1}{2}}(\mathbf{e}_w-\epsilon_w)\|^2\right]^{\frac{1}{2}}.
\end{aligned}$$

That is,

$$\|\mathbf{e}-\epsilon\|^2 \leq \sum_w \|Q_w^{\frac{1}{2}}(\mathbf{e}_w-\epsilon_w)\|^2.$$

Therefore, based on (5.7), the quasi-uniformity of $\{w\}$, and inequality (5.4), one gets

$$\begin{aligned}
\|\mathbf{e}-\epsilon\|^2 &\leq \sum_w \|Q_w^{\frac{1}{2}}(\mathbf{e}_w-\epsilon_w)\|^2 \leq \sum_w \|\mathbf{e}_w-\epsilon_w\|^2 \\
&\leq \delta \sum_w \frac{\mathbf{e}_w^T S_w \mathbf{e}_w}{\|S_w\|} \leq \frac{\delta}{\eta \|A\|^2} \sum_w \mathbf{e}_w^T S_w \mathbf{e}_w \\
&\leq \frac{\delta}{\eta \|A\|^2}\, \mathbf{e}^T A^T A\mathbf{e} \leq \frac{\delta}{\eta \|A\|}\, \mathbf{e}^T A\mathbf{e}.
\end{aligned}$$

□

We will use estimate (5.6) to show that the two-grid method with the Richardson iteration matrix $M = \frac{\|A\|}{\omega} I$, $\omega \in (0, 2)$, which leads to $\widetilde{M} = M(2M - A)^{-1}M = \frac{\|A\|^2}{\omega^2}(2\frac{\|A\|}{\omega} I - A)^{-1}$, is uniformly convergent. More specifically, we have the following main spectral equivalence result.

Theorem 5.1. *The algebraic two-grid preconditioner, based on the Richardson smoother* $M = \frac{\|A\|}{\omega}\, I$, $\omega \in (0, 2)$, *and the coarse space based on* P *constructed by the window spectral AMG method, is spectrally equivalent to* A *and the following estimate holds,*

$$\mathbf{v}^T A\mathbf{v} \le \mathbf{v}^T B\mathbf{v} \le \frac{\delta}{\eta\omega(2 - \omega)}\ \mathbf{v}^T A\mathbf{v}.$$

The constants $\frac{\delta}{\eta}$ *is from the coarse-grid approximation property* (5.6).

Proof. One first notices that

$$\mathbf{w}^T \widetilde{M}\mathbf{w} = \frac{\|A\|^2}{\omega^2}\ \mathbf{w}^T \left(2\frac{\|A\|}{\omega} I - A\right)^{-1} \mathbf{w} \le \frac{\|A\|}{\omega(2 - \omega)}\ \mathbf{w}^T\mathbf{w} = \frac{1}{2 - \omega}\ \mathbf{w}^T M\mathbf{w}.$$

Then, based on the $\widetilde{M}$-norm minimization property of the projection $\pi_{\widetilde{M}}$, one has,

$$\begin{aligned}
((I - \pi_{\widetilde{M}})\mathbf{v})^T \widetilde{M}(I - \pi_{\widetilde{M}})\mathbf{v} &= \inf_{\epsilon \in \mathrm{Range}(P)} (\mathbf{v} - \epsilon)^T \widetilde{M}(\mathbf{v} - \epsilon)\\
&\le \frac{1}{2 - \omega} \inf_{\epsilon \in \mathrm{Range}(P)} (\mathbf{v} - \epsilon)^T M(\mathbf{v} - \epsilon)\\
&= \frac{\|A\|}{\omega(2 - \omega)} \inf_{\epsilon \in \mathrm{Range}(P)} \|\mathbf{v} - \epsilon\|^2\\
&\le \frac{1}{\omega(2 - \omega)}\ \frac{\delta}{\eta}\ \mathbf{v}^T A\mathbf{v}.
\end{aligned}$$

One can use the above estimate for $\mathbf{v} = (I - \pi_A)\mathbf{v}$. Since $(I - \pi_{\widetilde{M}})(I - \pi_A) = (I - \pi_{\widetilde{M}})$ the left-hand side of the above estimate does not change. Thus the corresponding two-grid preconditioner is spectrally equivalent to A with a constant

$$K = \sup_{\mathbf{v}} \frac{((I - \pi_{\widetilde{M}})\mathbf{v})^T \widetilde{M}(I - \pi_{\widetilde{M}})\mathbf{v}}{((I - \pi_A)\mathbf{v})^T A(I - \pi_A)\mathbf{v}} \le \frac{\delta}{\eta\, \omega(2 - \omega)}.$$

□

REFERENCES

1. O. Axelsson, "*Iterative Solution Methods*", Cambridge University Press, 1994.
2. O. Axelsson and I. Gustafsson, "*Preconditioning and two-level multigrid methods of arbitrary degree of approximation*", Mathematics of Computations, **40**(1983), pp. 219-242.
3. R. E. Bank and T. Dupont, "*Analysis of a two-level scheme for solving finite element equations*", Report CNA-159, Center for Numerical Analysis, The University of Texas at Austin, Austin, TX, 1980.
4. R. E. Bank, T. Dupont, and H. Yserentant, "*The hierarchical basis multigrid method*", Numerische Mathematik **52**(1988), pp. 427-458,

5. R. E. Bank, "*Hierarchical bases and the finite element method*", vol. 5 of Acta Numerica, Cambridge University Press, Cambridge, 1996, pp. 1–43.
6. D. Braess, "*The contraction number of a multigrid method for solving Poisson equation*", Numerische Mathematik **37**(1981), pp. 387-404.
7. A. Brandt, "*General highly accurate algebraic coarsening*", Electronic Transactions on Numerical Analysis **10**(2000), pp. 1–20.
8. W. L. Briggs, V. E. Henson, and S. F. McCormick, "*A Multigrid Tutorial*", 2nd edition, SIAM, 2000.
9. T. Chartier, R. Falgout, V. E. Henson, J. Jones, T. Manteuffel, S. McCormick, J. Ruge, and P. S. Vassilevski, "*Spectral AMGe (ρAMGe)*", SIAM Journal on Scientific Computing **25**(2003), pp. 1–26.
10. V. L. Eijkhout and P. S. Vassilevski, "*The role of the C.B.S. inequality in multilevel methods*", SIAM Review **33**(1991), pp. 405-419.
11. R. D. Falgout and P. S. Vassilevski, "*On Generalizing the AMG Framework*", SIAM Journal on Numerical Analysis, to appear.
12. J. W. Ruge and K. Stüben, "*Algebraic multigrid (AMG)*", in Multigrid Methods, S. F. McCormick, ed., vol. 3 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1987, pp. 73–130.
13. P. S. Vassilevski, "*On two ways of stabilizing the HB multilevel methods*", SIAM Review **39**(1997), pp. 18–53.
14. J. Xu and L. T. Zikatanov, "*The method of alternating projections and the method of subspace corrections in Hilbert space*", Journal of the American Mathematical Society **15**(2002) pp. 573–597.
15. H. Yserentant, "*On the multilevel splitting of finite element spaces*", Numerische Mathematik **49**(1986), pp. 379-412.